

The AWGN Red Alert Problem

Bobak Nazer, *Member, IEEE*, Yanina Shkel, *Graduate Student Member, IEEE*, and Stark C. Draper, *Member, IEEE*

Abstract—Consider the following unequal error protection scenario. One special message, dubbed the “red alert” message, is required to have an extremely small probability of missed detection. The remainder of the messages must keep their average probability of error and probability of false alarm below a certain threshold. The goal then is to design a codebook that maximizes the error exponent of the red alert message while ensuring that the average probability of error and probability of false alarm go to zero as the blocklength goes to infinity. This red alert exponent has previously been characterized for discrete memoryless channels. This paper completely characterizes the optimal red alert exponent for additive white Gaussian noise channels with block power constraints.

I. INTRODUCTION

Communication networks are increasingly being taxed by the enormous demand for instantly available, streaming multimedia. Ideally, we would like to maximize the reliability and data rate of a system while simultaneously minimizing the delay. Yet, in the classical fixed blocklength setting, the reliability function of a code goes to zero as the rate approaches capacity even in the presence of feedback. This seems to imply that, close to capacity, it is impossible to keep delay low and reliability high. However, this lesson is partially an artifact of the block coding framework. The achievable tradeoff changes in a streaming setting where all bits do not need to be decoded by a fixed deadline, but rather, each individual bit must be recovered after a certain delay. In this setting, the reliability function measures how quickly the error probability on each bit estimate decays as a function of the delay. Surprisingly, the achievable error exponent can be quite large at capacity if a noiseless feedback link is available and cleverly exploited [1], [2].

The distinguishing feature of these streaming architectures with feedback is the use of an ultra-reliable special codeword that is transmitted to notify the decoder when it is about to make an error. While this “red alert” codeword requires a significant fraction of the decoding space to attain its very large error exponent, the remaining “standard” codewords merely need their error probability to vanish in the blocklength. One question that seems intimately connected to the

streaming delay-reliability tradeoff is how large the red alert error exponent can be made for a fixed blocklength codebook of a given rate. Beyond this streaming motivation, the red alert problem is also connected to certain sensor network scenarios. For example, consider a setting where sensors must send regular updates to the basestation using as little power as possible, i.e., using the standard codewords. If an anomaly is detected, the sensors are permitted to transmit at higher power in order to alert the basestation with high reliability, which corresponds to our red alert problem.

Prior work has characterized the red alert exponent for discrete memoryless channels (DMCs) [3]–[5]. In this paper, we determine the red alert exponent for point-to-point additive white Gaussian noise (AWGN) channels that operate under block power constraints on both the regular and red alert messages. We derive matching upper and lower bounds on the red alert exponent with a focus on the resulting high-dimensional geometry of the decoding regions. Our code construction can be viewed as a generalization of that used in the discrete case.

A. Related Work

Previous studies on protecting a special message over a DMC have relied on some variant of the following code construction. First, designate the special codeword to be the repetition of a particular input symbol. Then, generate a fixed composition codebook at the desired rate. This composition is chosen to place the “standard” codewords as far as possible from the special codeword (as measured by the Kullback-Leibler (KL) divergence between induced output distributions) while still allocating each codeword a decoding region large enough to ensure a vanishing probability of error. By construction, the rest of the space is given to the special codeword. Early work by Kudryashov used this strategy to achieve very high error exponents in the bit error setting under an expected delay constraint [1].

In [3], Borade, Nakiboğlu, and Zheng study “bit”-wise and “message”-wise unequal error protection (UEP) problems and error exponents. The red alert problem is a message-wise UEP problem in which one message is special and the remaining messages are standard. While [3] focuses on general DMCs near capacity, Lemma 1 of that paper develops a general sharp bound on the red alert exponent for DMCs at any rate below capacity (both with and without feedback). Specializing to the exponent achieved at capacity, let \mathcal{X} denote the input alphabet, $\{p_{Y|X}(\cdot|x)\}_{x \in \mathcal{X}}$ the channel transition matrix, and $p_Y^*(\cdot)$ the capacity-achieving output distribution of the DMC. Then, the optimal red alert exponent at capacity is

$$E_{\text{ALERT}}(C) = \max_{x \in \mathcal{X}} D(p_Y^*(\cdot) \| p_{Y|X}(\cdot|x)) \quad (1)$$

B. Nazer, Y. Shkel, and S. C. Draper were supported by the Air Force Office of Scientific Research under Grant FA9550-09-1-0140 and by the National Science Foundation under grant CAREER 0844539. The material in this paper was presented in part at the 48th Annual Allerton Conference on Communication, Control, and Computing, Monticello, IL, September 2010 and at the 6th Annual Information Theory and Applications Workshop, UCSD, La Jolla, CA, February 2011.

B. Nazer was with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI, 53706, USA and is now with the Department of Electrical and Computer Engineering, Boston University, Boston, MA, 02215 (email: bobak@bu.edu). Y. Shkel and S. C. Draper are with the Department of Electrical and Computer Engineering, University of Wisconsin, Madison, WI, 53706, USA; (ysshkel@wisc.edu; sdraper@ece.wisc.edu)

where $D(\cdot\|\cdot)$ is the KL divergence. We also mention recent work by Nakiboğlu *et al.* [5], [6] that considers the generalization where a strictly positive error exponent is required of the standard messages.

For the binary symmetric channel (BSC), the optimal red alert exponent has a simple and illustrative form. This exponent can be inferred from the general expression in [3, Lemma 1] or via a direct proof due to Sahai and Draper [4] (which appeared concurrently with the conference version [7] of [3]). Let p denote the crossover probability of the BSC and q the probability that a symbol in the codebook is a one. Then, the optimal red alert exponent as a function of rate $R < C$ for the BSC is

$$E_{\text{ALERT}}(R) = \max_{h_B(q*p) - h_B(p) \leq R} D(q * p \| p) \quad (2)$$

where $h_B(p) = -p \log p - (1-p) \log(1-p)$, $q * p = p(1-q) + q(1-p)$, and $D(\tilde{p} \| p) = \tilde{p} \log \left(\frac{\tilde{p}}{p} \right) + (1-\tilde{p}) \log \left(\frac{1-\tilde{p}}{1-p} \right)$.

Csiszár studied a related problem where multiple special messages require higher reliability in [8]. Upper bounds for multiple special messages with different priority levels were also developed in [3]. In [9], Borade and Sanghavi examined the red alert problem from a coding theoretic perspective. As shown by Wang [10], similar issues arise in certain sparse communication problems where the receiver must determine whether a codeword was sent or the transmitter was silent.

The fundamental mechanism through which high red alert exponents are achieved is a binary hypothesis test. By designing the induced distributions at the output of the channel to be far apart as measured by KL divergence, we can distinguish whether the red alert or some standard codeword was sent. The test threshold is biased to minimize the probability of missed detection and is analyzed via an application of Stein's Lemma. This sort of biased hypothesis test occurs in numerous other communication settings with feedback, such as [11]–[13] and, as mentioned earlier, these codes are also used as a component in streaming data systems (see, for instance, [1], [2], [14], [15]). There is also a rich literature on the interplay between hypothesis testing and information theory, which we cannot do justice to here (see, for instance, [16]–[18]).

II. PROBLEM STATEMENT

First, we mention some of our notational choices. We will use boldface lowercase letters to denote column vectors, $\mathbf{0}$ to denote the all zeros vector, and $\mathbf{1}$ to denote the all ones vector. Throughout the paper, the log function is taken to be the natural logarithm and rate is measured in nats instead of bits. We use $\|\mathbf{x}\|$ to denote the Euclidean norm of the vector \mathbf{x} .

Definition 1 (Messages): The transmitter has a message $w \in \{0, 1, 2, \dots, M\}$ that it wants to convey to the receiver. One of the messages, $w = 0$, is a red alert message that will be afforded extra error protection. We assume the red alert message is chosen with some probability greater than 0 and the remaining messages are chosen with equal probability.

Definition 2 (Encoder): The encoder \mathcal{E} maps the message w into a length- n real-valued codeword \mathbf{x} for transmission over the channel, $\mathcal{E} : \{0, 1, 2, \dots, M\} \rightarrow \mathbb{R}^n$. Let $\mathbf{x}(w)$

denote the codeword used for message w and let \mathcal{C} denote the entire codebook, $\mathcal{C} = \{\mathbf{x}(0), \mathbf{x}(1), \dots, \mathbf{x}(M)\}$. The codebook must satisfy both an average block power constraint across codewords,

$$\frac{1}{M} \sum_{w=1}^M \|\mathbf{x}(w)\|^2 \leq n P_{\text{avg}}. \quad (3)$$

In addition, the red alert codeword must satisfy a less stringent power constraint,

$$\|\mathbf{x}(0)\|^2 \leq n P_{\text{alert}}, \quad (4)$$

for some $P_{\text{alert}} \geq P_{\text{avg}}$. The *rate* of the codebook is

$$R = \frac{1}{n} \log M \quad (5)$$

nats per channel use.

Remark 1: Note that our codebook average power constraint (3) is less strict than the usual block power constraint $\|\mathbf{x}(w)\|^2 \leq n P_{\text{avg}}$. Our achievable scheme can be easily modified to meet this constraint using expurgation. Furthermore, our red alert power constraint (4) is less strict than a peak power constraint $|x_i(0)|^2 \leq P_{\text{alert}} \forall i$, where $x_i(0)$ denotes the i th symbol of the red alert codeword. Our scheme sets the red alert codeword to be $\mathbf{x}(0) = -\sqrt{P_{\text{alert}}} \mathbf{1}$, which naturally satisfies a peak power constraint. Therefore, our main results hold under an average power constraint and peak power constraint as well.

Remark 2: We omit the red alert codeword from the average block power constraint for the sake of simplicity. Another possibility would be to consider only an average block power constraint over both the standard and red alert codeword. This would lead to two different tensions between maximizing the red alert exponent and maximizing the rate. The first would be the allocation of the decoding regions and the second would be the allocation of power based on the probability of a red alert message. By using two separate power constraints, we can state our results in a simpler form that does not depend on the red alert probability.

Definition 3 (Channel): The channel outputs the transmitted vector, corrupted by independent and identically distributed (i.i.d.) Gaussian noise:

$$\mathbf{y} = \mathbf{x} + \mathbf{z} \quad (6)$$

where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, N \mathbf{I}^{n \times n})$ for some noise variance $N > 0$.

Definition 4 (Decoder): The signal observed by the receiver is sent into a decoder which produces an estimate \hat{w} of the transmitted message w , $\mathcal{D} : \mathbb{R}^n \rightarrow \{0, 1, 2, \dots, M\}$.

Definition 5 (Error Probability): We are concerned with three quantities, the *probability of missed detection* of the red alert message p_{MD} , the *probability of false alarm* p_{FA} , and the *average probability of error* of all other messages p_{MSG} :

$$p_{\text{MD}} = \mathbb{P}(\hat{w} \neq w | w = 0) \quad (7)$$

$$p_{\text{FA}} = \mathbb{P}(\hat{w} = 0 | w \neq 0) \quad (8)$$

$$p_{\text{MSG}} = \frac{1}{M} \sum_{w=1}^M \mathbb{P}(\hat{w} \neq w | w \neq 0, \hat{w} \neq 0). \quad (9)$$

Definition 6 (Error Exponent): We say that a red alert exponent of $E_{\text{ALERT}}(R)$ is achievable if for every $\epsilon > 0$ and n

large enough, there exists a rate R encoder and a decoder such that

$$-\frac{1}{n} \log(p_{\text{MD}}) > E_{\text{ALERT}}(R) - \epsilon \quad (10)$$

$$p_{\text{FA}} < \epsilon \quad (11)$$

$$p_{\text{MSG}} < \epsilon. \quad (12)$$

In other words, we would like the red alert codeword to have as large an error exponent as possible while keeping the other error probabilities small. The standard codewords do not need to have a positive error exponent. Of course, the rate must be lower than the AWGN capacity, $R \leq C$, where

$$C \triangleq \frac{1}{2} \log \left(1 + \frac{P_{\text{avg}}}{N} \right). \quad (13)$$

A. High-Dimensional Geometry

We now review some basic facts of high-dimensional geometry that will be useful in our analysis.

Let $\mathcal{B}_n(\mathbf{a}, r)$ denote the n -dimensional ball centered at $\mathbf{a} \in \mathbb{R}^n$ with radius $r > 0$. Recall that the volume of $\mathcal{B}_n(\mathbf{a}, r)$ is

$$\text{Vol}(\mathcal{B}_n(\mathbf{a}, r)) = \frac{r^n \pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} \quad (14)$$

where $\Gamma(\cdot)$ is the gamma function [19, Ch. 1, Eq. 16]. We define $\mathcal{S}_n(\mathbf{a}, r)$ to be the surface of $\mathcal{B}_n(\mathbf{a}, r)$. Its surface area (or, more precisely, the $(n-1)$ -dimensional volume of its surface) is

$$\text{Vol}(\mathcal{S}_n(\mathbf{a}, r)) = \frac{nr^{n-1} \pi^{n/2}}{\Gamma(\frac{n}{2} + 1)} \quad (15)$$

[19, Ch. 1, Eq. 19]. The dimension of the $\text{Vol}(\cdot)$ function will always be clear from the context. We also define

$$\mathcal{T}_n(\mathbf{a}, r_1, r_2) \triangleq \{\mathbf{x} : r_1 \leq \|\mathbf{x} - \mathbf{a}\| \leq r_2\} \quad (16)$$

to be the spherical shell centered at \mathbf{a} from radius r_1 to r_2 .

The angle between two n -dimensional vectors \mathbf{a} and \mathbf{b} is

$$\angle(\mathbf{a}, \mathbf{b}) = \cos^{-1} \left(\frac{\mathbf{a}^T \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} \right) \quad (17)$$

where $\cos^{-1}(\cdot)$ takes values between 0 and π . Let $\mathcal{V}_n(\mathbf{a}, \mathbf{b}, \theta)$ denote the n -dimensional cone with its origin at \mathbf{a} , its center axis running from \mathbf{a} to \mathbf{b} , and of half-angle θ which takes values from 0 to $\pi/2$. The solid angle $\Omega(\theta)$ of an n -dimensional cone of half-angle θ is the fraction of surface area that it carves out of an n -dimensional sphere,

$$\Omega(\theta) = \frac{\text{Vol}(\mathcal{V}_n(\mathbf{0}, \mathbf{1}, \theta) \cap \mathcal{S}_n(\mathbf{0}, r))}{\text{Vol}(\mathcal{S}_n(\mathbf{0}, r))}. \quad (18)$$

Note that the solid angle is the same for any sphere radius $r > 0$.

Lemma 1 (Shannon): The solid angle of a cone with half-angle θ satisfies

$$\Omega(\theta) = \frac{\sin^n \theta}{\sqrt{2\pi n} \sin \theta \cos \theta} \left(1 + O\left(\frac{1}{n}\right) \right).$$

See the math leading up to Equation 28 in [20] for a proof.

III. MAIN RESULT

In the binary case, the simplest characterization of the optimal codebook is a statistical one: the red alert codeword is the zero vector and the remaining codewords are of a constant composition. From one perspective, this can be visualized as placing the red alert codeword in the “center” of the space with the other codewords encircling it (see Figure 1). This corresponds to choosing the red alert codeword to be the all zeros (or all ones) vector. The standard codewords are generated using the distribution that maximizes the KL divergence between output distributions while still supporting a rate R . While this two-dimensional illustration is quite useful for understanding the binary case, it can be misleading in the Gaussian case. Specifically, it suggests that we should place the red alert codeword at the origin which turns out to be suboptimal.

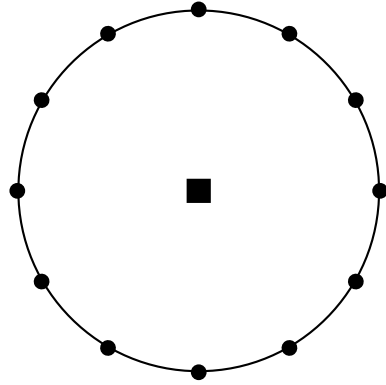


Fig. 1. For a BSC, we can visualize the red alert codeword (solid square) sitting in the “center” of the codebook and the standard codewords (solid circles) occupying a thin shell around it. While this illustration is generally sufficient for developing the intuition behind the discrete case, it does not capture the full story in the Gaussian case.

Another way of looking at the binary construction is to visualize each fixed composition as a parallel (or circle of constant latitude) on a sphere (see Figure 2). That is, the code lives on the Hamming cube in n dimensions, which can be imagined as a sphere by taking the all zeros and all ones vectors as the two poles and specifying the parallels by their Hamming weight. From this viewpoint, the binary construction sets the red alert codeword to be one of the poles and chooses the remaining codewords on the furthest parallel that can support a codebook of rate R . This perspective leads naturally to the right construction for the Gaussian case. Essentially, the standard codewords are placed uniformly along a constant parallel. This can be achieved by generating the standard codewords using a capacity-achieving code with a fraction α of the total power. The red alert codeword is placed at the furthest limit of the red alert power constraint (e.g., at $-\sqrt{P_{\text{alert}}}\mathbf{1}$) and the standard codewords are offset in the opposite direction (e.g., by $\sqrt{(1-\alpha)P_{\text{avg}}}\mathbf{1}$). See Figure 3 for an illustration. In the high-dimensional limit, most of the codewords will live on a parallel, thus mimicking the binary construction. This scheme leads us to the optimal red alert exponent.

Theorem 1: For an AWGN channel with red alert power

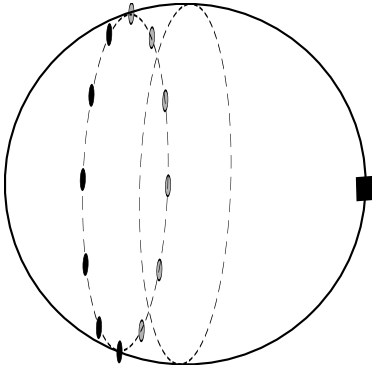


Fig. 2. From an alternate viewpoint, we can visualize the red alert codeword (solid square) sitting on the pole of a sphere and the standard codewords (solid circles) on a parallel on the opposite hemisphere. Our code construction for the Gaussian case is inspired by this picture. If the red alert power constraint is larger than the average power constraint, the red alert codeword should be placed on the sphere's axis but off its the surface, directly above the pole.

constraint P_{alert} , average power constraint P_{avg} , and rate R , the optimal red alert exponent is

$$E(R) = \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R}))}}{2N} - R.$$

We prove achievability in Lemma 7 and provide a matching upper bound in Lemma 11.

In the conference version of this paper [21], we used a different code construction that lead to a smaller achievable red alert exponent. The codewords were generated uniformly on the sphere of radius $\sqrt{nP_{\text{avg}}}$ and we only kept those that fell within a cone of appropriate half-angle. This type of construction turns out not to achieve as dense a packing as the construction used in this paper. In Appendix D, we explore the reasons why this occurs in the binary case. In Appendix E, we state the achievable red alert exponent for the conical construction.

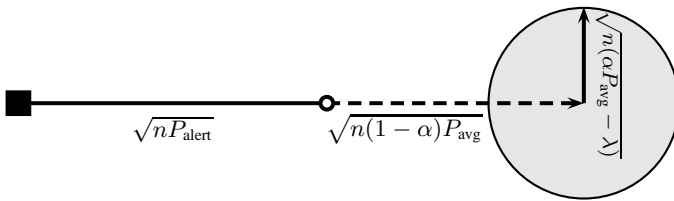


Fig. 3. Red alert codebook construction. The red alert codeword (solid square) is placed at $-\sqrt{P_{\text{alert}}}\mathbf{1}$ which takes it a distance $\sqrt{nP_{\text{alert}}}$ from the origin (circle). The standard codewords (shaded region) are drawn i.i.d. according to a Gaussian distribution with variance $\alpha P_{\text{avg}} - \lambda$. These codewords are pushed away from the origin by an offset $\sqrt{(1 - \alpha)P_{\text{avg}}}\mathbf{1}$ (dashed line).

IV. CODEBOOK CONSTRUCTION

Our codebook construction for \mathcal{C} consists of the following steps:

- 1) Choose $\epsilon > 0$ so that $R < C - \epsilon$.
- 2) The red alert codeword is placed at the boundary of the red alert power constraint, $\mathbf{x}(0) = -\sqrt{P_{\text{alert}}}\mathbf{1}$.

- 3) Choose $0 < \alpha \leq 1$ so that

$$R + \epsilon = \frac{1}{2} \log \left(1 + \frac{\alpha P_{\text{avg}}}{N} \right) \quad (19)$$

and choose $\lambda > 0$ so that

$$R + \frac{\epsilon}{2} = \frac{1}{2} \log \left(1 + \frac{\alpha P_{\text{avg}} - \lambda}{N} \right). \quad (20)$$

- 4) Draw e^{nR} codewords $\mathbf{v}(1), \dots, \mathbf{v}(e^{nR})$ i.i.d. according to a Gaussian distribution with mean zero and variance $\alpha P_{\text{avg}} - \lambda$.
- 5) To each of these codewords, add an offset $\sqrt{(1 - \alpha)P_{\text{avg}}}\mathbf{1}$ so that the transmitted codeword for each message (other than $w = 0$) is $\mathbf{x}(w) = \sqrt{(1 - \alpha)P_{\text{avg}}}\mathbf{1} + \mathbf{v}(w)$.

We will show that this procedure yields a random codebook \mathcal{C} whose false alarm probability and average probability of error are both less than ϵ . Afterwards, we will characterize the probability of missed detection for the red alert codeword. This will in turn imply the existence of a good fixed codebook.

V. ACHIEVABILITY

In this section, we will show that the red alert error exponent stated in Theorem 1 is achievable. We begin by stating useful large deviations bounds that will play a role in both the proof of the achievability and of the converse. Next, we show that any standard codeword plus noise lies at a certain distance from the red alert codeword with high probability. Afterwards, we argue that, with high probability, any standard codeword plus noise is contained in a cone of a certain half-angle that is centered on the red alert codeword. By combining the distance and angle bounds, we can constrain the decoding region for the standard codewords to the intersection of a cone with a shell. The remainder of \mathbb{R}^n can thus be allocated to the decoding region for the red alert codeword, for which we will bound the resulting probability of a missed detection.

A. Large Deviations Bounds

Our upper and lower bounds on the probability of error are proven by deriving bounds on the size and shape of the decoding regions and then applying Cramér's Theorem to get large deviations bounds. Define $g_X(a)$ to be the moment generating function of a random variable X ,

$$g_X(a) \triangleq \mathbb{E}[e^{aX}], \quad (21)$$

and $I_X(b)$ to be the Fenchel-Legendre transform [22, Definition 2.2.2] of $\log(g_X(\cdot))$,

$$I_X(b) \triangleq \sup_a [ab - \log(g_X(a))]. \quad (22)$$

Theorem 2 (Cramér): Let $S_n = \frac{1}{n} \sum_i X_i$ be the normalized sum of n i.i.d. variables X_1, \dots, X_n with finite mean and rate function $I_X(b)$. Then, for every closed subset $\mathcal{F} \subset \mathbb{R}$,

$$\mathbb{P}(S_n \in \mathcal{F}) \leq 2 \exp \left(-n \inf_{b \in \mathcal{F}} I_X(b) \right), \quad (23)$$

and, for every open subset $\mathcal{G} \subset \mathbb{R}$,

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n \in \mathcal{G}) \geq - \inf_{b \in \mathcal{G}} I_X(b). \quad (24)$$

See, for instance, [22, Theorem 2.2.3] for a proof. We will be particularly interested in how this bound applies to the length of i.i.d. Gaussian vectors, which corresponds to setting the X_i to be Chi-square random variables (with one degree of freedom). The moment generating function for such random variables is $g_X(a) = \frac{1}{\sqrt{1-2a}}$ which yields a rate function of $I_X(b) = \frac{1}{2}(b - 1 - \log b)$.

B. Distance Bounds

The following lemma formalizes the notion that the squared ℓ_2 -norm of an i.i.d. Gaussian vector concentrates sharply around its variance. Thus, for large n , the decoding region can be restricted to a thin spherical shell.

Lemma 2: Let \mathbf{z} be a length- n vector with i.i.d. zero-mean Gaussian entries of variance N . Then, for any $\beta > 0$,

$$\mathbb{P}(\|\mathbf{z}\|^2 \geq nN(1 + \beta)) \leq 2 \exp\left(-\frac{n}{2}(\beta - \log(1 + \beta))\right)$$

and, for any $0 < \beta < 1$,

$$\mathbb{P}(\|\mathbf{z}\|^2 \leq nN(1 - \beta)) \leq 2 \exp\left(-\frac{n}{2}(-\beta - \log(1 - \beta))\right).$$

See Appendix A for the proof.

Recall that the Q-function returns the probability that a scalar Gaussian random variable with mean zero and unit variance is greater than or equal to $t > 0$,

$$Q(t) \triangleq \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp\left(-\frac{t^2}{2}\right) dx, \quad (25)$$

and is upper bounded as

$$Q(t) < \frac{1}{2} \exp\left(-\frac{t^2}{2}\right).$$

The next lemma is about the well-known fact that an i.i.d. Gaussian vector is approximately orthogonal to any fixed vector.

Lemma 3: Let \mathbf{z} be a length- n vector with i.i.d. zero-mean Gaussian entries with variance N and let \mathbf{a} be a length- n vector with $\|\mathbf{a}\|^2 = n\alpha$ for some fixed $\alpha > 0$. Then, for any $\delta > 0$ and n large enough,

$$\mathbb{P}(|\mathbf{a}^T \mathbf{z}| \geq \delta \|\mathbf{a}\|^2) < \delta. \quad (26)$$

See Appendix A for the proof.

In Figure 4, the codebook is illustrated from the perspective of the origin. Using the above lemma, it can be shown that all but a vanishing fraction of codewords have power close to P_{avg} and are nearly orthogonal with respect to any fixed vector. We now characterize how far away a codeword plus noise is from the red alert codeword with high probability.

Lemma 4: For any $\delta > 0$ and n large enough, the distance from the red alert codeword to the codeword for a standard message, $w \in \{1, 2, \dots, e^{nR}\}$, plus noise is at least L with high probability,

$$\mathbb{P}(\|\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z}\| \geq L) > 1 - \delta \quad (27)$$

$$L = \sqrt{n(P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}(1 - \alpha)P_{\text{avg}}} - \lambda - \delta)}.$$

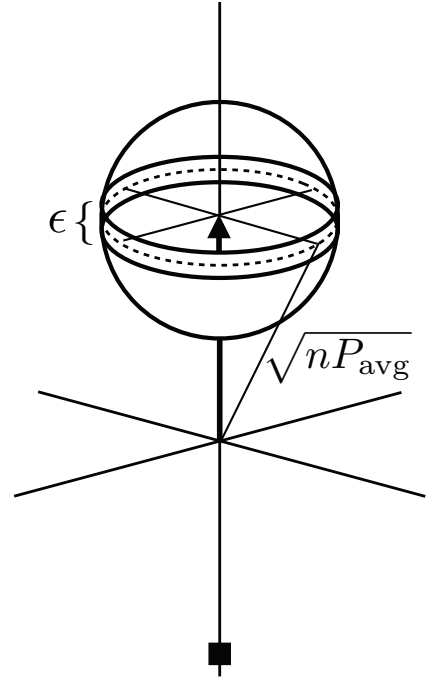


Fig. 4. From the perspective of the origin, most of the codewords are concentrated in an ϵ -shell of power $\alpha P_{\text{avg}} - \lambda$ that is offset away from the origin with power $(1 - \alpha)P_{\text{avg}}$. Thus, with high probability, any random codeword meets the power constraint.

See Appendix A for the proof.

C. Angle Bounds

We now upper bound the n -dimensional angle between a fixed vector and the same vector plus i.i.d. Gaussian noise.

Lemma 5: Let \mathbf{z} be a length- n vector with i.i.d. zero-mean Gaussian entries with variance N and let \mathbf{a} be a length- n vector with $\|\mathbf{a}\|^2 = n\alpha$ for some fixed $\alpha > 0$. For any $\delta > 0$ and n large enough, the probability that the angle between \mathbf{a} and $\mathbf{a} + \mathbf{z}$ exceeds $\cos^{-1}\left(\sqrt{\frac{\alpha}{\alpha + N}}\right) + \delta$ is upper bounded by δ ,

$$\mathbb{P}\left(\angle(\mathbf{a}, \mathbf{a} + \mathbf{z}) \geq \cos^{-1}\left(\sqrt{\frac{\alpha}{\alpha + N}}\right) + \delta\right) < \delta. \quad (28)$$

See Appendix B for the proof.

In Figure 5, we have depicted the distance L and the angle ψ from the red alert codeword to a standard codeword plus noise. Notice that both the noise and the codewords are (nearly) orthogonal to the axis along which the red alert codeword lies.

Now consider a cone centered on the red alert codeword that contains a standard codeword plus noise with high probability. The next lemma upper bounds the required half-angle for the cone.

Lemma 6: Let $\mathcal{V}_n(\mathbf{x}(0), \mathbf{0}, \psi)$ denote the cone centered on the red alert codeword with axis running towards the origin and half-angle ψ . For any $\delta > 0$, $w \in \{1, 2, \dots, e^{nR}\}$, and n

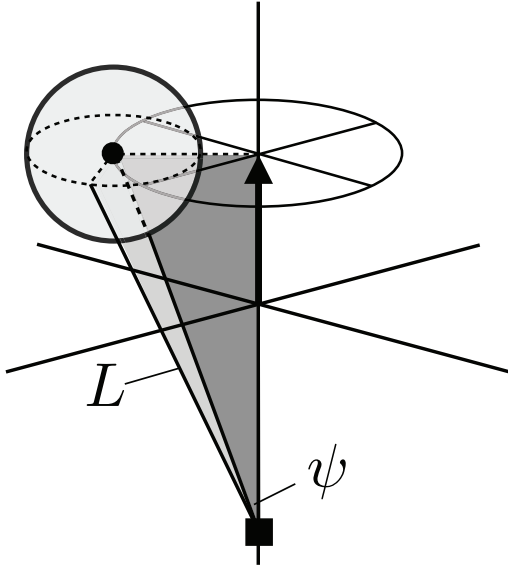


Fig. 5. With high probability, a Gaussian codeword, a Gaussian noise vector, and the red alert codeword vector are all nearly orthogonal to each other. Conditioning on this event, we can derive the minimum distance L and the angle ψ between a codeword plus noise and the red alert codeword.

large enough, if the half-angle ψ is greater than or equal to

$$\sin^{-1} \left(\sqrt{\frac{\alpha P_{\text{avg}} + N - \lambda}{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}} + N - \lambda}} \right) + \delta$$

then the cone contains the codeword for message w plus noise with high probability, i.e.,

$$\mathbb{P}(\mathbf{x}(w) + \mathbf{z} \in \mathcal{V}_n(\mathbf{x}(0), \mathbf{0}, \psi)) > 1 - \delta. \quad (29)$$

See Appendix B for the proof.

D. Red Alert Exponent

Now that we know the decoding region can be confined to a conical shell, we can bound the probability of missed detection for the red alert codeword.

Lemma 7: For any rate R , the following red alert exponent is achievable

$$E(R) = \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R}))}}{2N} - R.$$

Proof: Choose $\delta > 0$. In Lemma 4, L is a lower bound on the distance between the red alert codeword and a standard codeword plus noise. From Lemma 6, we have an upper bound on the half-angle needed to capture a standard codeword plus noise in the cone $\mathcal{V}_n(\mathbf{x}(0), \mathbf{0}, \psi)$ centered on the red alert codeword. If the received vector lies in the cone and is at least distance L from the red alert codeword, then the decoder assumes the red alert message was not transmitted. Otherwise, it declares that the red alert message was sent. For n large enough, we know that the probability that a random codeword plus noise, $\mathbf{x}(w) + \mathbf{z}$, leaves this region is at most ϵ . Therefore, the probability of false alarm (averaged over the randomness in the codebook) is upper bounded by ϵ .

If the received vector falls in the decoding region for standard messages, we simply subtract the offset $\sqrt{(1-\alpha)P_{\text{avg}}}$ and apply a maximum likelihood decoder to make an estimate of the transmitted message. Since the rate of the codebook is chosen to be slightly less than the capacity (for the power level $\alpha P_{\text{avg}} - \lambda$), it is straightforward to show that the average probability of error for a given message is at most ϵ .

Since the average false alarm probability and average error probability are small, it follows that there exists at least one fixed codebook with a small false alarm probability and average error probability. We now turn to upper bounding the probability of missed detection. Assume the red alert codeword is transmitted. Define

$$\beta = \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}} - \lambda - \delta}{N}. \quad (30)$$

where λ is specified by step 3) of the codebook construction in Section IV. Using Lemma 2, the probability that the noise pushes the red alert codeword further than L (as specified in Lemma 4) can be upper bounded by

$$\mathbb{P}(\|\mathbf{x}(0) + \mathbf{z}\| > L) = \mathbb{P}(\|\mathbf{x}(0) + \mathbf{z}\|^2 > nN(1 + \beta)) \quad (31)$$

$$\leq \exp\left(-\frac{n}{2}(\beta - \log(1 + \beta))\right). \quad (32)$$

The probability that the received vector falls into the cone of half-angle ψ is given by the fraction of surface area of a sphere carved out by the cone. Using Lemma 1, this can be calculated as

$$\begin{aligned} & \mathbb{P}(\mathbf{x}(0) + \mathbf{z} \in \mathcal{V}_n(\mathbf{x}(0), \mathbf{0}, \psi)) \\ &= \frac{\sin^n \psi}{\sqrt{2\pi n} \sin \psi \cos \psi} \left(1 + O\left(\frac{1}{n}\right)\right). \end{aligned} \quad (33)$$

Pulling terms into the exponent we get

$$\begin{aligned} & \exp\left(-n\left(-\log(\sin \psi) + \frac{1}{n} \log(\sqrt{2\pi n} \sin \psi \cos \psi)\right.\right. \\ & \quad \left.\left.+ O(1/n)\right)\right). \end{aligned} \quad (34)$$

For n large enough, we get that the probability is upper bounded by $\exp\left(-n(-\log(\sin \psi) - \delta)\right)$.

Since the noise is an i.i.d. Gaussian vector, its magnitude and direction are independent. Therefore, the probability of missed detection is upper bounded as

$$\begin{aligned} P_{MD} &= \mathbb{P}(\|\mathbf{x}(0) + \mathbf{z}\| > L) \mathbb{P}(\mathbf{x}(0) + \mathbf{z} \in \mathcal{V}_n(\mathbf{x}(0), \mathbf{0}, \psi)) \\ &\leq \exp\left(-\frac{n}{2}(\beta - \log(1 + \beta) - 2\log(\sin \psi) - 2\delta)\right) \end{aligned}$$

for n large enough. For λ and δ small enough and n large enough, the exponent $\frac{\beta}{2} - \frac{1}{2}\log(1 + \beta) - \log(\sin \psi) - \delta$ can

be made equal to

$$\begin{aligned}
& \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}}}{2N} \\
& - \frac{1}{2} \log \left(\frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}} + N}}{N} \right) \\
& + \frac{1}{2} \log \left(\frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}} + N}}{\alpha P_{\text{avg}} + N} \right) - \epsilon \\
& = \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}}}{2N} \\
& - \frac{1}{2} \log \left(1 + \frac{\alpha P_{\text{avg}}}{N} \right) - \epsilon \\
& = \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}}}{2N} - R - \epsilon.
\end{aligned}$$

Finally, we can solve for α in terms of R to get $\alpha = \frac{N}{P_{\text{avg}}}(e^{2R} - 1)$. Substituting this into the expression above yields the desired result. \blacksquare

Note that at $R = 0$, the coherent gain $2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R}))} = 2\sqrt{P_{\text{alert}}P_{\text{avg}}}$, which is the largest benefit we could hope for. At $R = C$, the coherent gain vanishes.

Remark 3: We can interpret our achievability result from a hypothesis testing perspective. Let \mathcal{H}_0 denote the event that a standard codeword is transmitted and let \mathcal{H}_1 denote the event that the red alert codeword is transmitted. Under \mathcal{H}_0 , the entries of \mathbf{y} are i.i.d. according to a Gaussian distribution with mean $\sqrt{(1-\alpha)P_{\text{avg}}}$ and variance $\alpha P_{\text{avg}} + N$. Under \mathcal{H}_1 , the entries are i.i.d. Gaussian with mean $-\sqrt{P_{\text{alert}}}$ and variance N . Using the Chernoff-Stein Lemma [23, Theorem 11.8.3], we can bound the missed detection probability of the optimal hypothesis test via the KL divergence between the two distributions, $D(\mathcal{N}(\sqrt{(1-\alpha)P_{\text{avg}}}, \alpha P_{\text{avg}} + N) \parallel \mathcal{N}(-\sqrt{P_{\text{alert}}}, N))$. A bit of calculation will reveal that this KL divergence corresponds exactly to the red alert exponent. One can obtain the same exponent by plugging these distributions into the red alert exponent expression from [3, Lemma 1]. However, this does not in itself constitute a proof as the results of [3] are for DMCs without cost constraints.

VI. CONVERSE

We now develop an upper bound on the red alert exponent. Our bound relies on the fact that, in order to recover the standard messages reliably, we must allocate a significant volume of the output space for decoding them, which contributes to the probability of missed detection. An overview of the main steps in the proof is provided below.

- In Lemma 8, we argue that a constant fraction of the codewords live in a thin shell and strictly satisfy the power and error constraints.
- With high probability, the standard codewords plus noise are concentrated in a thin shell. Lemma 9 establishes this fact as well as the minimum volume required for the decoding region to attain a given probability of error.
- To minimize the probability of missed detection, we should pack this volume into the thin shell to maximize

the distance from the red alert codeword (see Figure 6 for an illustration). Lemma 10 bounds the distance and angle from the red alert codeword to the resulting decoding region (see Figure 7 for an illustration).

- Finally, in Lemma 11, we bound the probability that the noise carries the red alert codeword into the decoding region for the standard codewords.

Lemma 8: Assume that a sequence of codebooks satisfies the average block power constraint P_{avg} and has average probability of error p_{MSG} that tends to zero. Then for any $\gamma > 0$ and n large enough, there exists a shell of width γ that contains $e^{n(R-\gamma)}$ codewords, each with probability of error at most $(2/\gamma)p_{\text{MSG}}$, and average power at most $P_{\text{avg}}(1-\gamma)^{-1}$. See Appendix C for the proof.

Lemma 9: Assume that, for some $\gamma, \rho > 0$, $e^{n(R-\gamma)}$ codewords, each with probability of error at most $(2/\gamma)p_{\text{MSG}}$ lie in the shell $\mathcal{T}_n(\mathbf{0}, \sqrt{n\rho}, \sqrt{n\rho} + \gamma)$. Then, for n large enough, the decoding region for these codewords must include a subset of the noise-inflated shell $\mathcal{T}_n(\mathbf{0}, \sqrt{n(\rho + N - \gamma)}, \sqrt{n(\rho + N + \gamma)})$ with volume at least

$$V_{\text{MIN}} = e^{n(R-\gamma)} \frac{(n\pi(N - \gamma))^{n/2}}{\Gamma(\frac{n}{2} + 1)}.$$

See Appendix C for the proof.

Lemma 10: Assume that a sequence of codebooks has rate R and an average probability of error p_{MSG} that tends to zero as n increases. Then, for sufficiently small ϵ and n large enough, the probability of missed detection p_{MD} is lower bounded by the probability that the noise vector has squared norm $\|\mathbf{z}\|^2$ between $L^2 + n\epsilon$ and $L^2 + 2n\epsilon$ and lies at an angle $\angle(\mathbf{z}, \mathbf{1})$ between $\psi(1 - \epsilon)$ and $\psi(1 + 2\epsilon)$ where

$$\begin{aligned}
L^2 &= n \left(P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R}))} \right) \\
\psi &= \sqrt{\frac{Ne^{2R}}{P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R}))}}}.
\end{aligned}$$

Proof: Consider the standard codewords from a red alert codebook. From Lemma 8, for any $\gamma > 0$ and n large enough, at least $e^{n(R-\gamma)}$ codewords with power at most $P_{\text{avg}}(1-\gamma)^{-1}$ and probability of error at most $(2/\gamma)p_{\text{MSG}}$ must lie in a shell $\mathcal{T}_n(\mathbf{0}, \sqrt{n\rho}, \sqrt{n\rho} + \gamma)$ for some $\rho > 0$. From Lemma 9, it follows that the decoding region for these codewords falls within the noise-inflated shell $\mathcal{T}_n(\mathbf{0}, \sqrt{n(\rho + N - \gamma)}, \sqrt{n(\rho + N + \gamma)})$ and has volume at least V_{MIN} .

To get our lower bound, we need to pack this volume in the noise-inflated shell such that it minimizes p_{MD} . Since the noise vector is i.i.d. Gaussian, the probability that the red alert codeword is pushed to a certain point is determined solely by a decreasing function of the distance. Let \mathcal{G}_d denote the set of all points at distance d or greater from the red alert codeword

$$\mathcal{G}_d = \{\mathbf{z} : \|\mathbf{z} - \mathbf{x}(0)\| \geq d\}. \quad (35)$$

The optimal volume packing corresponds to the intersection of the set \mathcal{G}_d and the noise shell $\mathcal{T}_n(\mathbf{0}, \sqrt{n(\rho + N - \nu)}, \sqrt{n(\rho + N + \nu)})$ with d chosen such

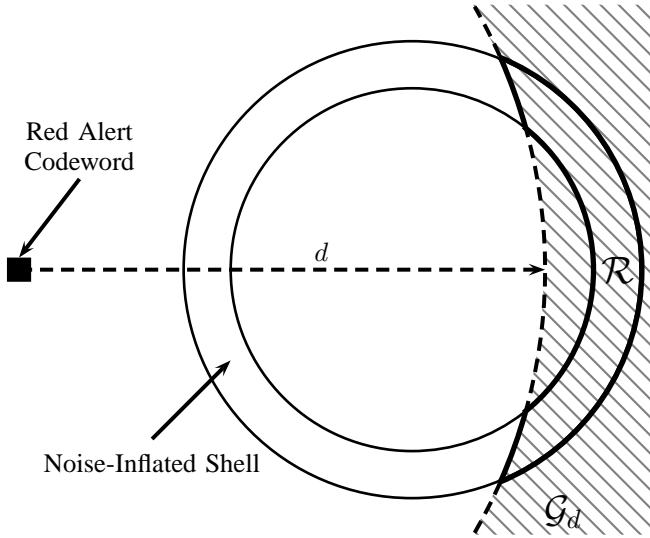


Fig. 6. To attain the desired probability of error, the decoding region for the standard codewords must include a subset of the noise-inflated shell with volume at least V_{MIN} . To minimize the probability of missed detection, we place this volume as far from the red alert codeword (square) as possible. Let \mathcal{G}_d denote the set of points at distance d or greater from the red alert codeword. The decoding region \mathcal{R} is the intersection of \mathcal{G}_d and the shell, where d is chosen to capture volume V_{MIN} .

that the volume of the set is equal to V_{MIN} . Let \mathcal{R} denote the resulting region and see Figure 6 for an illustration.

Let $\mathcal{R}_{\text{EDGE}}$ denote the set of points in \mathcal{R} that sit at the minimum distance to the red alert codeword,

$$\mathcal{R}_{\text{EDGE}} = \left\{ \mathbf{u} \in \mathcal{R} : \|\mathbf{u} - \mathbf{x}(0)\| = \min_{\mathbf{w} \in \mathcal{R}} \|\mathbf{w} - \mathbf{x}(0)\| \right\},$$

and let $\mathbf{v}^* \in \mathcal{R}_{\text{EDGE}}$ be any of these points. Let L^* and ψ^* denote the distance and angle from the red alert codeword $\mathbf{x}(0)$ to \mathbf{v}^* . We now seek to bound these quantities through a bound on the angle from the origin to \mathbf{v}^* .

Let θ^* denote the half-angle of a cone, centered on the origin that contains the region \mathcal{R} (and thus includes \mathbf{v}^*). The volume of this cone must be at least equal to that of \mathcal{R} since \mathcal{R} is a subset of the noise shell. Therefore, θ^* is lower bounded by the half-angle θ of a cone whose volume is equal to the volume of \mathcal{R} (see Figure 7 for an illustration). Combining (14) and Lemma 1, for n large enough, the volume of this cone is upper bounded by

$$\frac{(n\pi(\rho + N + \nu)\sin^2\theta)^{n/2}}{\Gamma(\frac{n}{2} + 1)}. \quad (36)$$

Now, since we require this quantity to exceed V_{MIN} , we can lower bound θ by

$$\theta \geq \sin^{-1} \left(e^{R-\nu} \sqrt{\frac{N-\nu}{\rho + N + \nu}} \right). \quad (37)$$

We can further lower bound θ by setting ρ to its maximum value $P_{\text{avg}}(1-\gamma)^{-1}$,

$$\theta \geq \sin^{-1} \left(e^{R-\nu} \sqrt{\frac{N-\nu}{P_{\text{avg}}(1-\gamma)^{-1} + N + \nu}} \right). \quad (38)$$

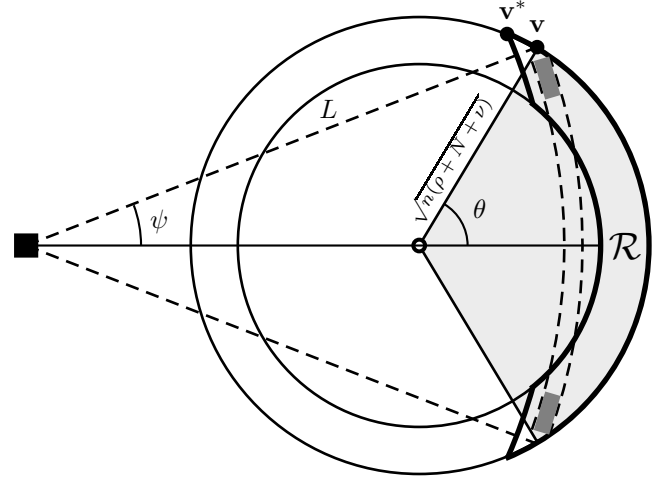


Fig. 7. Illustration of successive lower bounds on the probability of missed detection, p_{MD} . The red alert codeword is represented by a square and the origin by a circle. The decoding region \mathcal{R} is denoted by a thick line. We would like to characterize the distance L^* and angle ψ^* to the edge of \mathcal{R} , represented by the point \mathbf{v}^* . To do so, we consider a cone with half-angle θ (shaded region) with the same volume as \mathcal{R} . The intersection of this cone with the outer surface of \mathcal{R} contains a point \mathbf{v} at distance $L > L^*$ and angle $\psi < \psi^*$. In our final lower bound, we only consider the event that the received vector lies in the subset of \mathcal{R} at distance slightly larger than L and within an angle between $\psi - \epsilon$ to $\psi + \epsilon$ from the red alert codeword (illustrated by dark patches).

Thus, for any $\delta > 0$, ν and γ small enough, and n large enough, θ^* is lower bounded as follows

$$\theta^* \geq \theta \geq \sin^{-1} \left(e^{R-\delta} \sqrt{\frac{N}{P_{\text{avg}} + N}} \right). \quad (39)$$

The distance L^* from $\mathbf{x}(0)$ to \mathbf{v}^* is upper bounded by the distance L to a point \mathbf{v} that lies on the intersection of the outer shell (at distance $\sqrt{n(P_{\text{avg}}(1-\gamma)^{-1} + N + \nu)}$ from the origin) and the cone of half-angle θ . Without loss of generality, assume that the red alert codeword is placed at $\mathbf{x}(0) = -\mu\mathbf{1}$ for some $\mu > 0$.¹ The direction of the red alert codeword is not important since we will always fill the noise shell relative to this direction. Then L^2 is at least

$$\begin{aligned} & \left(\sqrt{n\mu} + \cos\theta \sqrt{n(P_{\text{avg}}(1-\gamma)^{-1} + N + \nu)} \right)^2 \\ & + \left(\sin\theta \sqrt{n(P_{\text{avg}}(1-\gamma)^{-1} + N + \nu)} \right)^2. \end{aligned} \quad (40)$$

For any $\delta > 0$, γ and ν small enough, and n large enough, this quantity is itself upper bounded by

$$n \left(\mu + P_{\text{avg}} + N + 2\cos\theta \sqrt{\mu(P_{\text{avg}} + N)} + \delta \right).$$

The half-angle ψ of a cone, centered on the red alert codeword, that contains the point \mathbf{v} is lower bounded by

$$\sin^{-1} \left(\frac{\sin\theta \sqrt{n(P_{\text{avg}}(1-\gamma)^{-1} + N + \nu)}}{\sqrt{n(\mu + P_{\text{avg}} + N + 2\cos\theta \sqrt{\mu(P_{\text{avg}} + N)} + \delta)}} \right)$$

¹It is straightforward to prove that placing the red alert codeword exactly at the origin is suboptimal.

which, for any $\eta > 0$, γ , ν , and δ small enough, and n large enough is itself lower bounded by

$$\sin^{-1} \left(\frac{(1-\eta) \sin \theta \sqrt{P_{\text{avg}} + N}}{\sqrt{\mu + P_{\text{avg}} + N + 2 \cos \theta \sqrt{\mu(P_{\text{avg}} + N)}}} \right).$$

The probability of missed detection decreases if the distance L^* from $\mathbf{x}(0)$ to \mathbf{v}^* is increased. The angle ψ^* will simultaneously decrease. Thus, by setting $\mu = P_{\text{alert}}$, we further lower bound the probability of missed detection. Using the relation $\sin^2 \theta + \cos^2 \theta = 1$ combined with (39), we find that $\cos \theta \leq \sqrt{1 - e^{2R-2\delta} \frac{N}{N+P_{\text{avg}}}}$. Plugging in μ and θ , we obtain the following upper bound on $(L^*)^2$:

$$n(P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R-2\delta}))} + \delta)$$

and the following lower bound on ψ^* :

$$\sin^{-1} \sqrt{\frac{(1-\eta)^2 N e^{2R-2\delta}}{P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R-2\delta}))}}}$$

Finally, it follows that, for ϵ small enough (but greater than 0 for finite n) and n large enough, the optimal packing contains all points from squared distance $L^2 + n\epsilon$ to $L^2 + 2n\epsilon$ from the red alert codeword and angle $\psi(1-\epsilon)$ to $\psi(1-2\epsilon)$ where L and ψ are as in the statement of the theorem. Thus, the probability of missed detection is lower bounded by the event that the noise falls into this region. ■

Lemma 11: For any rate R , the red alert exponent is upper bounded by

$$E(R) \leq \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R}))}}{2N} - R.$$

Proof: Lemma 10 established that the probability of missed detection is lower bounded by the event that the noise has squared length between $L^2 + n\epsilon$ to $L^2 + 2n\epsilon$ and angle between $\psi(1-\epsilon)$ and $\psi(1-2\epsilon)$ for some ϵ that tends to 0 as n tends to infinity. We now lower bound the probability of this event. Define

$$\beta = \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R}))}}{N}. \quad (41)$$

Since the magnitude and angle of an i.i.d. Gaussian vector are independent, the probability of missed detection is lower bounded as follows:

$$p_{\text{MD}} \geq \mathbb{P}(L^2 + n\epsilon \leq \|\mathbf{z}\|^2 \leq L^2 + 2n\epsilon) \cdot \mathbb{P}(\mathbf{z} \in \{\mathcal{V}_n(\mathbf{0}, \mathbf{1}, \psi(1-\epsilon)) \setminus \mathcal{V}_n(\mathbf{0}, \mathbf{1}, \psi(1-2\epsilon))\})$$

By Lemma 1, for n large enough, the second term in the product can be lower bounded by

$$(\sin(\psi(1-\epsilon)^2))^n - (\sin(\psi(1-2\epsilon)))^n \quad (42)$$

which, for n large enough, is itself lower bounded by

$$((1-\epsilon)^3 \sin(\psi(1-\epsilon)))^n \quad (43)$$

$$= \exp\left(-n\left(-\log((1-\epsilon)^3 \sin(\psi(1-\epsilon)))\right)\right). \quad (44)$$

Now, substituting in the lower bound on ψ from Lemma 10, we arrive at the following lower bound

$$\exp\left(-n\left(-R + \frac{1}{2}\log(1+\beta) - 3\log(1-\epsilon)\right)\right). \quad (45)$$

Using the upper bound on L from Lemma 10 and applying Theorem 2 for Chi-square random variables (and noting that ϵ and ν go to zero as n goes to infinity), it follows that

$$\liminf_{n \rightarrow \infty} \left(-\frac{1}{n} \log \mathbb{P}(L^2 + n\epsilon \leq \|\mathbf{z}\|^2 \leq L^2 + 2n\epsilon) \right) \quad (46)$$

$$\geq \frac{\beta}{2} - \frac{1}{2} \log(1+\beta). \quad (47)$$

Combining this with the lower bound on the angle event in (45), the exponent of the probability of missed detection is lower bounded by

$$\liminf_{n \rightarrow \infty} \left(-\frac{1}{n} \log p_{\text{MD}} \right) \quad (48)$$

$$\geq \frac{\beta}{2} - \frac{1}{2} \log(1+\beta) + \frac{1}{2} \log(1+\beta) - R \quad (49)$$

$$= \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(P_{\text{avg}} + N(1 - e^{2R}))}}{2N} - R$$

as desired. ■

VII. PLOTS

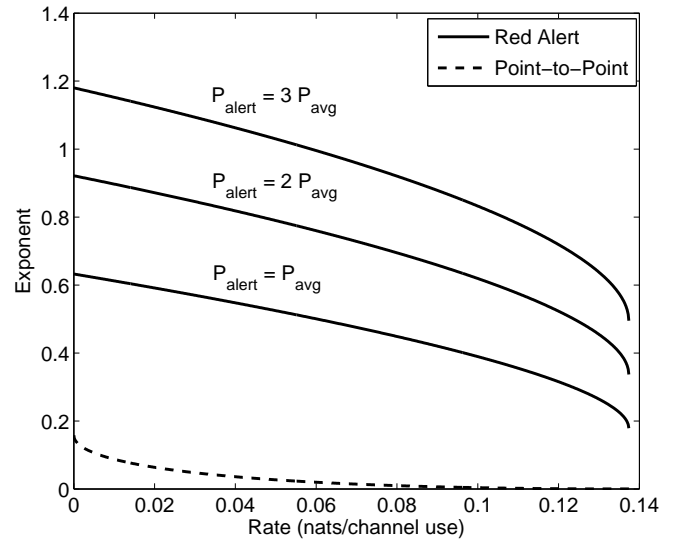


Fig. 8. Optimal red alert exponent for $P_{\text{avg}} = -5\text{dB}$ with $P_{\text{alert}} = P_{\text{avg}}, 2P_{\text{avg}}, 3P_{\text{avg}}$. An upper bound on the point-to-point AWGN error exponent is provided for comparison.

In Figures 8 and 9, we have plotted the optimal red alert exponent for $P_{\text{avg}} = -5\text{dB}$ and $P_{\text{avg}} = 15\text{dB}$, respectively, with red alert power constraints $P_{\text{alert}} = P_{\text{avg}}, 2P_{\text{avg}},$ and $3P_{\text{avg}}$. For comparison, we have also plotted an upper bound on the AWGN point-to-point error exponent from [20, Equation 4]. Notice that the red alert exponent can be quite large at capacity, even when the red alert power constraint is equal to the average power constraint.

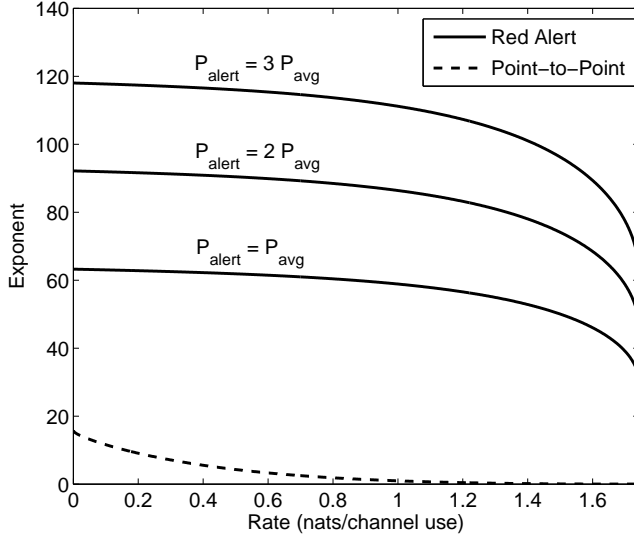


Fig. 9. Optimal red alert exponent for $P_{\text{avg}} = 15\text{dB}$ with $P_{\text{alert}} = P_{\text{avg}}, 2P_{\text{avg}}, 3P_{\text{avg}}$. An upper bound on the point-to-point AWGN error exponent is provided for comparison.

VIII. CONCLUSIONS

We have developed sharp bounds on the error exponent for distinguishing a single special message from 2^{nR} standard messages over an AWGN channel. As discussed in the introduction, these bounds can be used to characterize the performance of certain data streaming architectures, where each bit must be decoded after a given delay. An interesting question for future study is how well a single special message can be protected at a given finite blocklength, i.e., understanding the limits of unequal error protection in the non-asymptotic regime [24].

APPENDIX A

PROOFS FOR SECTION V-B

Proof of Lemma 2: The squared Euclidean distance is the sum of n i.i.d. squared Gaussian random variables with variance N . Therefore, $\frac{1}{nN}\|\mathbf{z}\|^2$ is the sum of n i.i.d. Chi-square random variables. Applying Theorem 2 and plugging in the Chi-square rate function of $I_Z(b) = \frac{1}{2}(b - 1 - \log b)$, it follows that

$$\mathbb{P}(\|\mathbf{z}\|^2 \geq nNb) \leq 2 \exp\left(-\frac{n}{2}(b - 1 - \log b)\right). \quad (50)$$

Substituting in $b = 1 + \beta$ yields the first bound and $b = 1 - \beta$ yields the second. ■

Proof of Lemma 3: First, we write the probability that $|\mathbf{a}^T \mathbf{z}|$ is greater than t in terms of the Q-function,

$$\mathbb{P}(|\mathbf{a}^T \mathbf{z}| \geq t) = 2Q\left(\frac{t}{\sqrt{N}\|\mathbf{a}\|}\right) \quad (51)$$

$$< \exp\left(-\frac{t^2}{2N\|\mathbf{a}\|^2}\right). \quad (52)$$

Substituting $t = \delta\|\mathbf{a}\|^2$ yields,

$$\mathbb{P}(|\mathbf{a}^T \mathbf{z}| \geq \delta\|\mathbf{a}\|^2) < \exp\left(-\frac{\delta^2 n \alpha}{2N}\right), \quad (53)$$

which can be driven arbitrarily close to zero for n large enough. ■

Proof of Lemma 4: We simply wish to bound the length of the vector from the special codeword to a standard codeword plus noise, $-\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z}$. By expanding terms, we obtain:

$$\|-\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z}\|^2 \quad (54)$$

$$= \left\| \left(\sqrt{P_{\text{alert}}} + \sqrt{(1-\alpha)P_{\text{avg}}} \right) \mathbf{1} + \mathbf{v}(w) + \mathbf{z} \right\|^2 \quad (55)$$

$$= \left(\sqrt{P_{\text{alert}}} + \sqrt{(1-\alpha)P_{\text{avg}}} \right)^2 \mathbf{1}^T \mathbf{1} + 2 \left(\sqrt{P_{\text{alert}}} + \sqrt{(1-\alpha)P_{\text{avg}}} \right) \mathbf{1}^T (\mathbf{v}(w) + \mathbf{z}) + \|\mathbf{v}(w) + \mathbf{z}\|^2. \quad (56)$$

The first term is $n(P_{\text{alert}} + (1-\alpha)P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}})$. The second term is the inner product of a fixed vector $2(\sqrt{P_{\text{alert}}} + \sqrt{(1-\alpha)P_{\text{avg}}})\mathbf{1}$ and an i.i.d. Gaussian vector $\mathbf{v}(w) + \mathbf{z}$ since $\mathbf{v}(w)$ is an element of a random Gaussian codebook. Thus, using Lemma 3, it can be shown that the probability that this inner product is less than $-n\delta/2$ is at most $\delta/2$ for n large enough. The third term is the squared norm of an i.i.d. Gaussian vector with mean zero and variance $\alpha P_{\text{avg}} + N - \lambda$. From Lemma 2, it follows that $\|\mathbf{v}(w) + \mathbf{z}\|^2$ is less than $n(\alpha P_{\text{avg}} + N - \lambda - \delta/2)$ with probability at most $\delta/2$ for n large enough. Combining these three bounds completes the proof. ■

APPENDIX B

PROOFS FOR SECTION V-C

Proof of Lemma 5: The angle between \mathbf{a} and $\mathbf{a} + \mathbf{z}$ is

$$\angle(\mathbf{a}, \mathbf{a} + \mathbf{z}) = \cos^{-1} \left(\frac{\mathbf{a}^T (\mathbf{a} + \mathbf{z})}{\|\mathbf{a}\| \|\mathbf{a} + \mathbf{z}\|} \right). \quad (57)$$

From Lemma 3, for any $\nu > 0$ and n large enough, the probability that $\mathbf{a}^T \mathbf{z} > \nu \|\mathbf{a}\|^2$ is at most ν . Therefore, since $\|\mathbf{a}\| = n\alpha$ we have that $\mathbf{a}^T (\mathbf{a} + \mathbf{z}) \geq (1 + \nu)n\alpha$ with probability at most ν . Combining Lemmas 2 and 3, we can also show that the probability that $\|\mathbf{a} + \mathbf{z}\| < (1 - \nu)\sqrt{n(\alpha + N)}$ is at most ν for n large enough. Thus, the probability that

$$\frac{\mathbf{a}^T (\mathbf{a} + \mathbf{z})}{\|\mathbf{a}\| \|\mathbf{a} + \mathbf{z}\|} < \frac{(1 + \nu)n\alpha}{\sqrt{n\alpha}(1 - \nu)\sqrt{n(\alpha + N)}} \quad (58)$$

$$= \frac{1 + \nu}{1 - \nu} \sqrt{\frac{\alpha}{\alpha + N}} \quad (59)$$

is at most 2ν . Choosing ν small enough yields the desired result. ■

Proof of Lemma 6: The angle between the axis of the cone and the standard codeword plus noise is

$$\angle(-\mathbf{x}(0), -\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z}) = \cos^{-1}(u) \quad (60)$$

$$u = \frac{(-\mathbf{x}(0))^T (-\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z})}{\|\mathbf{x}(0)\| \|\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z}\|}. \quad (61)$$

Since $\cos^{-1}(u)$ is a decreasing function of u , an upper bound on the angle can be obtained by lower bounding u . We will do

this by lower bounding the numerator and upper bounding the denominator (with high probability). Expanding the numerator yields:

$$(-\mathbf{x}(0))^T(-\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z}) \quad (62)$$

$$= \sqrt{P_{\text{alert}}} \left(\sqrt{P_{\text{alert}}} + \sqrt{(1-\alpha)P_{\text{avg}}} \right) \mathbf{1}^T \mathbf{1} \quad (63)$$

$$+ \sqrt{P_{\text{alert}}} \mathbf{1}^T (\mathbf{v}(w) + \mathbf{z}) . \quad (64)$$

The first term is simply $n\sqrt{P_{\text{alert}}}(\sqrt{P_{\text{alert}}} + \sqrt{(1-\alpha)P_{\text{avg}}})$. The second term is the inner product of a fixed vector and an i.i.d. Gaussian vector. Thus, using Lemma 3, it can be shown that for any $\nu > 0$ and n large enough, the probability that the second term is less than $-\nu n$ is at most ν . The denominator is composed of two terms. The first is simply $\|\mathbf{x}(0)\| = \sqrt{nP_{\text{alert}}}$. Following the proof of Lemma 4, it can be shown that the second term $\|-\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z}\|$ is greater than

$$\sqrt{n \left(P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}} - \lambda + \nu \right)}$$

with probability at most ν . Combining these bounds, we get that the probability that u is less than

$$\frac{\sqrt{P_{\text{alert}}} + \sqrt{(1-\alpha)P_{\text{avg}}} - \nu}{\sqrt{P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}} - \lambda + \nu}} \quad (65)$$

with probability at most 2ν . Thus, so long as the half-angle ψ is greater than or equal to

$$\cos^{-1} \left(\frac{\sqrt{P_{\text{alert}}} + \sqrt{(1-\alpha)P_{\text{avg}}}}{\sqrt{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}(1-\alpha)P_{\text{avg}}} + N - \lambda}} \right) + \delta$$

the cone contains $\mathbf{x}(w) + \mathbf{z}$ with probability at least $1 - \delta$ for n large enough. Applying the trigonometric identity $\sin^2(\psi) + \cos^2(\psi) = 1$ completes the proof. ■

APPENDIX C PROOFS FOR SECTION VI

Proof of Lemma 8: Observe that at least one codeword has power at most P_{avg} , otherwise the average will be larger than P_{avg} . If we remove this codeword's contribution from the average, the remaining codewords have average power at most $P_{\text{avg}} \frac{e^{nR}}{e^{nR}-1}$. Now, we can find a codeword whose power must be at most $P_{\text{avg}} \frac{e^{nR}}{e^{nR}-1}$. Removing this codeword yields an average of $P_{\text{avg}} \frac{e^{nR}}{e^{nR}-2}$. Continuing this process, we can remove γe^{nR} codewords that each have power at most $P_{\text{avg}}(1-\gamma)^{-1}$. By the same argument, we can find $(1-(\gamma/2))e^{nR}$ codewords that each have probability of error at most $(2/\gamma)p_{\text{MSG}}$. Therefore, at least $(\gamma/2)e^{nR}$ codewords must satisfy both these constraints simultaneously.

The selected codewords live in the sphere of radius $\sqrt{nP_{\text{avg}}(1-\gamma)^{-1}}$. We partition this sphere into shells of width γ each. It follows that at least one of these shells must contain $\frac{\gamma^2}{2\sqrt{n(P_{\text{avg}}(1-\gamma)^{-1})}} e^{nR}$ codewords. Finally, select n large enough so that $e^{n(R-\gamma)} \leq \frac{\gamma^2}{2\sqrt{n(P_{\text{avg}}(1-\gamma)^{-1})}} e^{nR}$. ■

Proof of Lemma 9: Assume that one of the codewords from the shell $\mathcal{T}_n(\mathbf{0}, \sqrt{n\rho}, \sqrt{n\rho} + \gamma)$ is transmitted. It follows

from Lemma 2, that for any $\gamma > 0$ and n large enough, the probability that $\|\mathbf{y}\|$ is larger than $\sqrt{n(\rho + N + \gamma)}$ or smaller than $\sqrt{n(\rho + N - \gamma)}$ is upper bounded by $p_{\text{MSG}} \frac{1}{\gamma}$. If the noise lands outside this “noise shell,” then we will assume that the transmitted codeword is decoded correctly. However, each codeword still needs to capture $1 - p_{\text{MSG}} \frac{3}{\gamma}$ probability inside the shell to ensure the error probability does not exceed $p_{\text{MSG}} \frac{2}{\gamma}$.

Now, consider the volume required for decoding a single codeword reliably. Since the noise is i.i.d. Gaussian, its probability distribution is rotationally invariant. This implies that the shape that uses the least volume to capture a given probability of error is a sphere centered on the codeword. Let $\sqrt{n\nu}$ be the radius of this sphere. By Lemma 2, if $\nu < N$, the probability that the noise falls inside this sphere goes to zero exponentially in n which implies the probability of error goes to one. Therefore, for n large enough, the probability of error will always exceed the desired probability of error (which is assumed to be bounded away from one). Using (14), we get that the decoding region of each codeword must have volume at least $\frac{(n\pi(N-\gamma))^{n/2}}{\Gamma(\frac{n}{2}+1)}$ for any $\gamma > 0$. We find that we will need a volume of at least

$$V_{\text{MIN}} = e^{n(R-\gamma)} \frac{(n\pi(N-\gamma))^{n/2}}{\Gamma(\frac{n}{2}+1)} \quad (66)$$

to reliably decode these codewords. ■

APPENDIX D OFFSET CODES VERSUS CONICAL CODES

We now develop some intuition for why the offset construction of Section IV is a better construction than the conical construction we used in our earlier work [21]. The difference between these two constructions is easier to understand in a discrete setting so we will analyze the corresponding constructions for a BSC with crossover probability p . For ease of analysis, we will calculate rate in bits per channel use (rather than nats per channel use).

First, recall that the BSC red alert exponent can be attained using a fixed composition codebook. Specifically, each of the 2^{nR} codewords is drawn independently and uniformly from the set of weight- nq binary sequences. If the rate is less than the induced mutual information, the average probability of error can be driven to zero

$$R < I(X; Y) = h_B(q * p) - h_B(p) . \quad (67)$$

The red alert codeword is taken to be all the all zeros vector. The decoder runs a hypothesis test between the two possible output distributions, Bernoulli(p) and Bernoulli($q * p$). The error exponent for the probability of missed detection is the KL divergence between the two distributions, $D(q * p \| p)$. As shown in [4], this is the optimal red alert exponent.

We can construct a *conical code* of parameter $q > 1/2$ by first drawing $2^{n(C-\epsilon)}$ codewords i.i.d. according to a Bernoulli($1/2$) distribution for some $\epsilon > 0$. Let \mathcal{C} denote the resulting set of codewords. To guarantee the same red alert exponent, we only keep those codewords with Hamming weight nq or greater and set the red alert codeword to be the

all zeros vector. We now bound the rate of this construction. Using Theorem 2, it can be shown that the probability that a Bernoulli(1/2) sequence \mathbf{x} has Hamming weight at least nq is upper bounded as

$$\mathbb{P}(\text{wt}(\mathbf{x}) \geq nq) \leq 2^{-nD(q\|0.5)}. \quad (68)$$

Take \mathcal{C}_q to be the set of subset of codewords in \mathcal{C} with Hamming weight nq or greater. Using (68), the expected size of \mathcal{C}_q is upper bounded by

$$\mathbb{E}[|\mathcal{C}_q|] \leq 2^{n(C-D(q\|0.5)-\epsilon)} \quad (69)$$

$$= 2^{n(1-h_B(p)-(1-h_B(q))-\epsilon)} \quad (70)$$

$$= 2^{n(h_B(q)-h_B(p)-\epsilon)}. \quad (71)$$

It can be shown with a Chernoff bound that the probability \mathcal{C}_q contains significantly more codewords vanishes doubly exponentially in n . Furthermore, it can be shown that the average probability of error of \mathcal{C}_q vanishes with n . Therefore, the rate of the conical codebook is $h_B(q) - h_B(p)$ for a red alert exponent of $D(q * p \| p)$.

Now, observe that $q < q * p < 1/2$ (unless either q or p is equal to $1/2$) so $h_B(q) < h_B(q * p)$, meaning that the rate of the conical construction is strictly less than the (optimal) constant composition construction. Intuitively, this means that the usual i.i.d. Bernoulli($\frac{1}{2}$) construction used to approach capacity does not pack codewords of higher (or lower) weights efficiently. Constraining the weight of codewords is essential to the hypothesis test that leads to the red alert exponent. The constant composition (or offset) construction is successful since it optimizes the packing of codewords of a given weight. A similar phenomenon occurs in the AWGN setting as shown below.

APPENDIX E

AWGN CONICAL CODES: RED ALERT EXPONENT

For completeness, we review the AWGN conical code that we proposed in [21] and the resulting red alert exponent. The construction is comprised of three main steps:

- 1) Place the red alert codeword at the limit of the red alert power constraint, $\mathbf{x}(0) = -\sqrt{P_{\text{alert}}}\mathbf{1}$.
- 2) Draw $2^{n(C-\epsilon)}$ codewords i.i.d. according to a Gaussian distribution with mean 0 and variance $P_{\text{avg}} - \epsilon$.
- 3) Of these codewords, only keep the first 2^{nR} that lie in the cone $\mathcal{V}_n(\mathbf{0}, \mathbf{1}, \theta + \epsilon)$ where $\theta = \sin^{-1}(e^{-(C-R)})$. (If there are fewer than 2^{nR} such codewords, declare an error.)

It can be shown that with high probability the resulting codebook contains 2^{nR} codewords inside the cone of half-angle θ . We now turn to bounding the distance and angle from the standard decoding region to the red alert codeword.

The distance can be bounded using the techniques used to prove Lemma 4. It follows that for any $\delta > 0$ and n large enough, the squared distance from the red alert codeword to a standard codeword plus noise is at least L with high probability,

$$\mathbb{P}(\|\mathbf{x}(0) + \mathbf{x}(w) + \mathbf{z}\| \geq L) > 1 - \delta \quad (72)$$

$$L^2 = n(P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}P_{\text{avg}}}\cos\theta + N - \delta). \quad (73)$$

Substituting in $\cos^2\theta = 1 - \sin^2\theta = 1 - e^{-2(C-R)}$, we get that L^2 is equal to

$$n(P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}P_{\text{avg}}(1 - e^{-2(C-R)})} - \delta).$$

Similarly, the techniques from Lemma 6 can be used to bound the angle. Let $\mathcal{V}_n(\mathbf{x}(0), \mathbf{0}, \psi)$ denote the cone centered on the red alert codeword with axis running towards the origin and half-angle ψ . For any $\delta > 0$ and n large enough, if the half-angle ψ is larger than

$$\sin^{-1}\left(\frac{e^{-(C-R)}\sqrt{P_{\text{avg}}}}{\sqrt{P_{\text{alert}} + P_{\text{avg}} + N + 2\sqrt{P_{\text{alert}}P_{\text{avg}}(1 - e^{-2(C-R)})}}}\right)$$

then the cone contains the codeword for message w plus noise with high probability, i.e.,

$$\mathbb{P}(\mathbf{x}(w) + \mathbf{z} \in \mathcal{V}_n(\mathbf{x}(0), \mathbf{0}, \psi)) > 1 - \delta. \quad (74)$$

Finally, these two bounds can be combined, as in the proof of Lemma 7, to get an achievable red alert exponent² of

$$E_{\text{ALERT}} = \frac{P_{\text{alert}} + P_{\text{avg}} + 2\sqrt{P_{\text{alert}}P_{\text{avg}}(1 - e^{-2(C-R)})}}{2N} + \frac{1}{2}\log\left(\frac{P_{\text{avg}} + N}{P_{\text{avg}}}\right) - R. \quad (75)$$

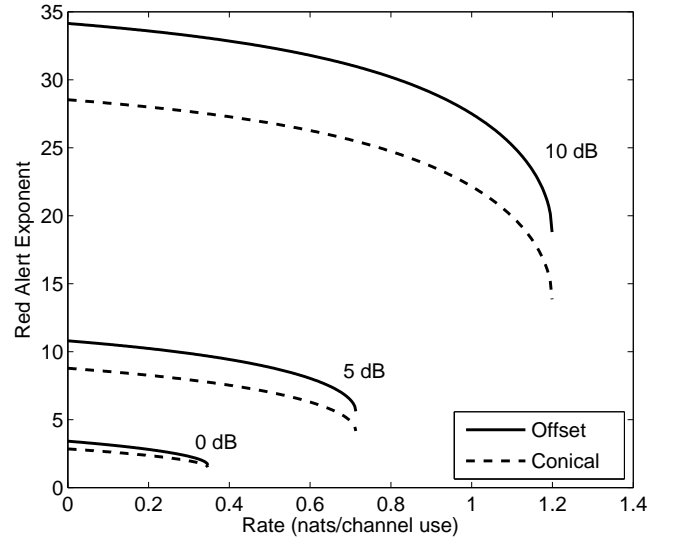


Fig. 10. Comparison of the red alert exponent attained by an (optimal) offset code construction and a conical code construction with an average power constraint of 0, 5, and 10dB with $P_{\text{alert}} = 2P_{\text{avg}}$.

In Figure 10, we have plotted this red alert exponent alongside the optimal one derived via the offset construction for average power constraints $P_{\text{avg}} = 0, 5$, and 10dB with $P_{\text{alert}} = 2P_{\text{avg}}$.

²Note that this is an improvement over the error exponent reported in Theorem 1 of [21] since we have used tighter upper bounds. Specifically, in [21], we did not completely take advantage of the fact that both the noise and the standard codewords are nearly orthogonal to any fixed vector and to each other with high probability.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers whose suggestions improved the presentation of this work. They would also like to thank B. Nakiboğlu for insightful discussions on the connections between this work and the discrete memoryless case.

REFERENCES

- [1] B. D. Kudryashov, "On message transmission over a discrete channel with noiseless feedback," *Problemy Peredachi Informatsii*, vol. 15, no. 1, pp. 3–13, 1979.
- [2] A. Sahai, "Why do block length and delay behave differently if feedback is present?," *IEEE Transactions on Information Theory*, vol. 54, pp. 1860–1886, May 2008.
- [3] S. Borade, B. Nakiboğlu, and L. Zheng, "Unequal error protection: An information-theoretic perspective," *IEEE Transactions on Information Theory*, vol. 55, pp. 5511–5539, December 2009.
- [4] A. Sahai and S. C. Draper, "The 'hallucination' bound for the BSC," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2008)*, (Toronto, Canada), July 2008.
- [5] B. Nakiboğlu, S. K. Gorantla, L. Zheng, and T. P. Coleman, "Bit-wise unequal error protection for variable length block codes with feedback," *IEEE Transactions on Information Theory*, Submitted January 2011. Available online: <http://arxiv.org/abs/1101.1934>.
- [6] S. K. Gorantla, B. Nakiboğlu, L. Zheng, and T. P. Coleman, "Bit-wise unequal error protection for variable length block codes with feedback," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2010)*, (Austin, TX), June 2010. Available online: <http://arxiv.org/abs/1101.1934>.
- [7] S. Borade, B. Nakiboğlu, and L. Zheng, "Some fundamental limits of unequal error protection," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2008)*, (Toronto, Canada), July 2008.
- [8] I. Csiszár, "Joint source-channel exponent," *Problems of Control and Information Theory*, vol. 9, no. 5, pp. 315–328, 1982.
- [9] S. Borade and S. Sanghavi, "Some fundamental coding theoretic limits of unequal error protection," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2009)*, June 2009.
- [10] D. Wang, "Distinguishing codes from noise: Fundamental limits and applications to sparse communication," Master's thesis, Massachusetts Institute of Technology, June 2012.
- [11] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Transactions on Information Theory*, vol. 9, pp. 136–143, July 1963.
- [12] G. D. Forney, Jr., "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Transactions on Information Theory*, vol. 14, pp. 206–220, March 1968.
- [13] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Problemy Peredachi Informatsii*, vol. 12, no. 4, pp. 10–30, 1976.
- [14] S. C. Draper and A. Sahai, "Noisy feedback improves communication reliability," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2006)*, (Seattle, WA), July 2006.
- [15] Y. Y. Shkel and S. C. Draper, "Cooperative reliability for streaming multiple access," in *Proceedings of the IEEE International Symposium on Information Theory (ISIT 2010)*, (Austin, TX), June 2010.
- [16] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Transactions on Information Theory*, vol. 20, pp. 405–417, July 1974.
- [17] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Transactions on Information Theory*, vol. 32, pp. 533–542, July 1986.
- [18] T. S. Han and K. Kobayashi, "The strong converse theorem for hypothesis testing," *IEEE Transactions on Information Theory*, vol. 35, pp. 178–180, January 1989.
- [19] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices, and groups*. New York, NY: Springer, 3rd ed., 1999.
- [20] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell System Technical Journal*, vol. 38, no. 3, pp. 611–656, 1959.
- [21] B. Nazer and S. C. Draper, "Gaussian red alert exponents: Geometry and code constructions," in *48th Annual Allerton Conference on Communications, Control, and Computing*, (Monticello, IL), September 2010.
- [22] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, NY: Springer, 1998.
- [23] T. Cover and J. Thomas, *Elements of Information Theory*. Hoboken, NJ: Wiley-Interscience, 2nd ed., 2006.
- [24] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, pp. 2307–2359, May 2010.